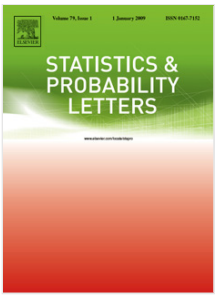
Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions

Célestin C. Kokonendji, Silvio S. Zocchi

Accepted Manuscript



Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions

Célestin C. Kokonendji, Silvio S. Zocchi

PII: S0167-7152(10)00202-6
DOI: 10.1016/j.spl.2010.07.008
Reference: STAPRO 5746

To appear in: *Statistics and Probability Letters*

Received date: 21 May 2010
Revised date: 11 July 2010
Accepted date: 12 July 2010

Please cite this article as: Kokonendji, C.C., Zocchi, S.S., Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters* (2010), doi:10.1016/j.spl.2010.07.008

Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions

Célestin C. Kokonendji [a,*],

[a] *University of Franche-Comté - LMB UMR 6623 CNRS - Besançon, France*

Silvio S. Zocchi [b].

[b] *University of São Paulo - ESALQ Piracicaba - SP, Brazil*

Abstract

Asymmetric discrete triangular distributions are introduced in order to extend the symmetric ones serving for discrete associated-kernels in the nonparametric estimation for discrete functions. The extension from one to two orders around the mode provides a large family of discrete distributions having a finite support. Establishing a bridge between Dirac and discrete uniform distributions, some different shapes are also obtained and their properties are investigated. In particular, the mean and variance are pointed out. Applications to discrete kernel estimators are given with solution to a boundary bias problem.

Key words: Asymmetric discrete distribution, discrete associated-kernel, finite support, limit distribution.

MSC 2010: Primary 62E15; Secondary 62G07.

1 Introduction

The recent notion of discrete associated-kernel for smoothing or estimating discrete functions requires a development of new families of discrete distributions. In this sense, Abdous and Kokonendji (2009) presented some asymptotic properties for discrete associated-kernel estimators of a *probability mass function* (pmf). Furthermore,

* *Corresponding address:* Université de Franche-Comté, UFR Sciences et Technique, Laboratoire de Mathématiques de Besançon - UMR 6623 CNRS, 16 route de Gray, 25030 Besançon cedex, France. Tel +33 381 666 341; Fax +33 381 666 623.

Email addresses: celestin.kokonendji@univ-fcomte.fr (Célestin C. Kokonendji), sszocchi@esalq.usp.br (Silvio S. Zocchi).

they pointed out the use of discrete associated-kernel from Dirac and symmetric discrete triangular distributions introduced by Kokonendji *et al.* (2007) and also from extensions of Dirac distribution proposed by Aitchison and Aitken (1976) and Wang and van Ryzin (1981). On the other hand, instead of pmf, one can use them to estimate discrete weighted functions (Kokonendji *et al.*, 2009a) or discrete regression functions (Kokonendji *et al.*, 2009b).

Let us firstly fix the definition and state in Theorem 1.2 some important properties presented in Kokonendji *et al.* (2007).

Definition 1.1 *Let $(m, a, h) \in \mathbb{Z} \times \mathbb{N} \times \mathbb{R}_+$. A denoted distribution $\mathcal{DT}(m; a; h)$ is said to be* ***symmetric discrete triangular*** *distribution with mode m, arm a and order h, if its pmf is*

$$f(y; m, a, h) = \frac{1}{(a+1)^{-h} D(a, h)} \left[1 - \frac{|y-m|^h}{(a+1)^h}\right], \quad y \in \{m, m \pm 1, \cdots, m \pm a\},$$

with

$$D(a, h) = (2a+1)(a+1)^h - 2\sum_{k=1}^{a} k^h.$$

Theorem 1.2 *Let Y be a random variable following $\mathcal{DT}(m; a; h)$. Then:*
(i) $\mathbb{E}(Y) = m$ and $\mathcal{DT}(m; a; h)$ is symmetric around m;
(ii) $Var(Y) = V(a, h)$ does not depend on m and is given by

$$V(a, h) = \frac{1}{D(a, h)} \left[\frac{a\,(2a+1)\,(a+1)^{h+1}}{3} - 2\sum_{k=1}^{a} k^{h+2}\right];$$

(iii) when $h \to 0$, $\mathcal{DT}(m; a; h)$ tends to $\mathcal{D}(m)$, i.e., the Dirac distribution at m;
(iv) when $h \to \infty$, $\mathcal{DT}(m; a; h)$ tends to $\mathcal{U}(\{m, m \pm 1, \cdots, m \pm a\})$, i.e., the discrete uniform distribution on the support $\{m, m \pm 1, \cdots, m \pm a\}$.

In this paper we mainly extend the symmetric discrete triangular distribution to a more general and flexible one, including asymmetry, keeping though the same support. In this sense, it provides a natural solution to the problem of boundary bias related to discrete associated-kernel estimators described by Kokonendji *et al.* (2007, 2009a); see also formulas (2) and Definition 5.1. For this reason, we must make sure that the mode is m, the expectation must tend to m when h goes to zero, and both properties (iii) and (iv) of Theorem 1.2 hold too.

Note finally that, although general discrete triangular distributions have attracted far less attention in the literature in contrast to the (continuous) triangular distributions (e.g. Johnson, 1997) for which the similar extensions are possible, they are now of such an interest that they were inserted in the dictionary of classical discrete distributions (e.g. Johnson *et al.*, 2005). The rest of this paper is organized as follows. Section 2 presents some details of the standard ($h = 1$) asymmetric discrete triangular distribution. Section 3 is devoted to the first extension of $\mathcal{DT}(m; a; h)$ following two arms and a single order. In Section 4 we conclude with the general discrete triangular distribution having both different two arms and two orders. In

Section 5 we present some applications to discrete kernel estimators with solutions to the problem of boundary bias.

2 Standard discrete triangular distribution

Let us first fix the notation related to the support of any discrete triangular distribution. We shall denote a given mode $m \in \mathbb{Z}$ and two arms $(a_1, a_2) \in \mathbb{N}^2$, by

$$\aleph_{m,a_1,a_2} = \aleph^*_{a_1,m} \cup \aleph_{m,a_2} = \aleph_{a_1,m} \cup \aleph^*_{m,a_2} = \aleph^*_{a_1,m} \cup \{m\} \cup \aleph^*_{m,a_2} \tag{1}$$

with $\aleph_{a_1,m} = \{m-k \ ; \ k = 0, 1, \cdots, a_1\}$, $\aleph^*_{a_1,m} = \aleph_{a_1,m} \setminus \{m\}$, $\aleph_{m,a_2} = \{m+k \ ; \ k = 0, 1, 2, \cdots, a_2\}$ and $\aleph^*_{m,a_2} = \aleph_{m,a_2} \setminus \{m\}$. Following the standard (continuous) triangular distribution (e.g. Johnson, 1997), we have the following definition in the discrete case.

Definition 2.1 *Let $(m, a_1, a_2) \in \mathbb{Z} \times \mathbb{N} \times \mathbb{N}$. A distribution $\mathcal{DT}(m; a_1, a_2)$ is said to be a **standard discrete triangular** distribution with mode m, left arm a_1 and right arm a_2, if its pmf is*

$$f(y; m, a_1, a_2) = \frac{1}{(a_1 + a_2 + 2)/2} \left[\left(1 - \frac{m-y}{a_1+1}\right) \mathbf{1}_{\aleph^*_{a_1,m}}(y) + \left(1 - \frac{y-m}{a_2+1}\right) \mathbf{1}_{\aleph_{m,a_2}}(y) \right],$$

where $\mathbf{1}_S(y)$ denotes the indicator function of any given set S that takes the value 1 for $y \in S$ and 0 otherwise.

Remark 2.2 *For $a_1 = a_2 = a$ we get the standard symmetric discrete triangular distribution from Definition 1.1 as $\mathcal{DT}(m; a, a) = \mathcal{DT}(m; a; h = 1)$.*

Figure 1 (a) presents some graphs of $\mathcal{DT}(m; a_1, a_2)$. We state the following proposition without proof.

Proposition 2.3 *Let $Y \sim \mathcal{DT}(m; a_1, a_2)$. Then*

$$\mathbb{E}(s^Y) = 2s^{m+1} \frac{(a_1+1)s^{a_2+1} + (a_2+1)s^{-(a_1+1)} - (a_1+a_2+2)}{(a_1+a_2+2)(a_1+1)^2(a_2+1)(s-1)}.$$

In particular, we have

$$\mathbb{E}(Y) = m + \frac{a_2 - a_1}{3},$$

$$Var(Y) = \mathbb{E}\left[Y - \mathbb{E}(Y)\right]^2 = \frac{1}{18}\left(a_1^2 + a_1a_2 + 3a_1 + 3a_2 + a_2^2\right),$$

$$\mathbb{E}\left[Y - \mathbb{E}(Y)\right]^3 = \frac{1}{270}(a_2 - a_1)(a_1 + 2a_2 + 3)(2a_1 + a_2 + 3) \quad \textit{and}$$

$$\mathbb{E}\left[Y - \mathbb{E}(Y)\right]^4 = \frac{1}{270}(a_1^2 + 3a_1 + a_1a_2 + 3a_2 + a_2^2)(2a_1^2 + 6a_1 + 2a_1a_2 + 6a_2 + 2a_2^2 - 3).$$

Since a_1 and a_2 are non-negative integer values, the sign of the skewness coefficient $(\mathbb{E}\left[Y - \mathbb{E}(Y)\right]^3 / Var^{3/2}(Y))$ will depend only on the sign of $(a_2 - a_1)$. Thus, the

distribution $\mathcal{DT}(m; a_1, a_2)$ will be skewed to the right if $a_2 > a_1$, skewed to the left if $a_2 < a_1$ and symmetric when $a_1 = a_2$. Furthermore, the kurtosis coefficient $(\mathbb{E}[Y - \mathbb{E}(Y)]^4 / Var^2(Y) - 3)$ will be always negative, i.e., $\mathcal{DT}(m; a_1, a_2)$ is always platykurtic.

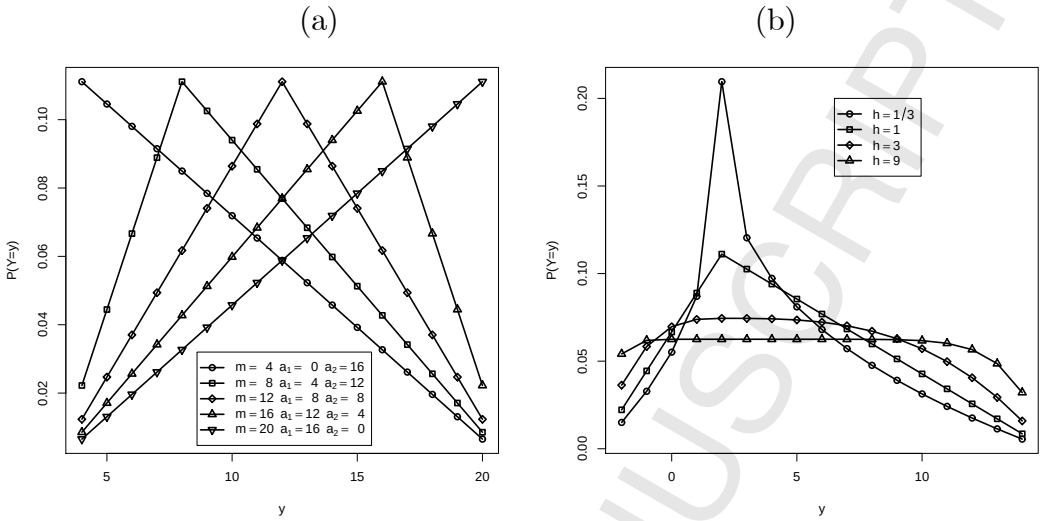


Fig. 1. (**a**) Standard discrete triangular distributions $\mathcal{DT}(m; a_1, a_2)$ for different values of m and arms such that $a_1 + a_2 = 16$ and (**b**) discrete (h-order) triangular distributions $\mathcal{DT}(m; a_1, a_2; h)$ with $m = 2$, $a_1 = 4$, $a_2 = 12$ and different values of h.

3 Discrete (h-order) triangular distribution

The asymmetric version of $\mathcal{DT}(m; a; h)$ is defined as follows.

Definition 3.1 *Let $(m, a_1, a_2, h) \in \mathbb{Z} \times \mathbb{N} \times \mathbb{N} \times \mathbb{R}_+$. A distribution $\mathcal{DT}(m; a_1, a_2; h)$ is said to be **discrete (h-order) triangular** distribution with mode m, left arm a_1, right arm a_2 and order h, if its pmf is*

$$f(y; m, a_1, a_2, h) = \frac{1}{D(a_1, a_2, h)} \left\{ \left[1 - \left(\frac{m - y}{a_1 + 1} \right)^h \right] \mathbf{1}_{\aleph^*_{a_1, m}}(y) + \left[1 - \left(\frac{y - m}{a_2 + 1} \right)^h \right] \mathbf{1}_{\aleph_{m, a_2}}(y) \right\},$$

with

$$D(a_1, a_2, h) = (a_1 + a_2 + 1) - (a_1 + 1)^{-h} \sum_{k=1}^{a_1} k^h - (a_2 + 1)^{-h} \sum_{k=1}^{a_2} k^h.$$

Remark 3.2 *(i) For $a_1 = a_2 = a$ we obtain all the symmetric discrete triangular distributions given in Definition 1.1 as $\mathcal{DT}(m; a, a; h) = \mathcal{DT}(m; a; h)$. (ii) For $h = 1$, $\mathcal{DT}(m; a_1, a_2; h = 1) = \mathcal{DT}(m; a_1, a_2)$ is the standard discrete distribution of Definition 2.1.*

Figure 1 (b) presents some graphs of $\mathcal{DT}(m; a_1, a_2; h)$ which can deform all the previous of Figure 1 (a). Some elementary properties from standard discrete triangular

are preserved but not elegant to be written (e.g. skewness, kurtosis and probability generating function). Here we show some results similar to the ones in Theorem 1.2.

Theorem 3.3 *Let $Y \sim \mathcal{DT}(m; a_1, a_2; h)$. Then:*
(i) $\mathbb{E}(Y) = m + A(a_1, a_2; h)$ *with*

$$A(a_1, a_2; h) = \frac{1}{D(a_1, a_2, h)} \left[\frac{a_2 (a_2 + 1)}{2} - \frac{a_1 (a_1 + 1)}{2} + \sum_{k=1}^{a_1} k \left(\frac{k}{a_1 + 1} \right)^h - \sum_{k=1}^{a_2} k \left(\frac{k}{a_2 + 1} \right)^h \right];$$

(ii) $Var(Y) = B(a_1, a_2; h) - [A(a_1, a_2; h)]^2$ *does not depend on* m *with*

$$B(a_1, a_2; h) = \frac{1}{D(a_1, a_2, h)} \left[\frac{a_2 (a_2 + 1) (2a_2 + 1)}{6} + \frac{a_1 (a_1 + 1) (2a_1 + 1)}{6} - \sum_{k=1}^{a_1} k^2 \left(\frac{k}{a_1 + 1} \right)^h - \sum_{k=1}^{a_2} k^2 \left(\frac{k}{a_2 + 1} \right)^h \right];$$

(iii) when $h \to 0$, $\mathcal{DT}(m; a_1, a_2; h)$ *tends to* $\mathcal{D}(m)$;
(iv) when $h \to \infty$, $\mathcal{DT}(m; a_1, a_2; h)$ *tends to* $\mathcal{U}(\aleph_{m,a_1,a_2})$.

PROOF. (i) : Using different decompositions of the support in (1) and Definition 3.1 with $D := D(a_1, a_2, h)$, one has successively

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y \in \aleph_{m,a_1,a_2}} y \; f(y; m, a_1, a_2, h) \\
&= \frac{1}{D} \left[\sum_{y \in \aleph_{m,a_1,a_2}} y - \sum_{y \in \aleph^*_{a_1 m}} y \left(\frac{m - y}{a_1 + 1} \right)^h - \sum_{y \in \aleph^*_{m,a_2}} y \left(\frac{y - m}{a_2 + 1} \right)^h \right] \\
&= \frac{1}{D} \left\{ mD + \sum_{k=1}^{a_1} \left[-k + k \left(\frac{k}{a_1 + 1} \right)^h \right] + \sum_{k=1}^{a_2} \left[k - k \left(\frac{k}{a_2 + 1} \right)^h \right] \right\}
\end{aligned}$$

which leads to the result.
(ii) : From (i) $\mathbb{E}(Y) = m + A$ with $A := A(a_1, a_2; h)$, one gets similarly

$$
\begin{aligned}
Var(Y) = & \sum_{y\in\aleph_{m,a_1,a_2}} (y-m-A)^2 \; f(y;m,a_1,a_2,h) \\
= & \frac{1}{D}\Bigg[\sum_{y\in\aleph_{m,a_1,a_2}} (y-m-A)^2 - \sum_{y\in\aleph^*_{a_1 m}} (y-m-A)^2 \left(\frac{m-y}{a_1+1}\right)^h \\
& - \sum_{y\in\aleph^*_{m,a_2}} (y-m-A)^2 \left(\frac{y-m}{a_2+1}\right)^h\Bigg] \\
= & \frac{1}{D}\Bigg\{A^2 + \sum_{k=1}^{a_1}\left[(k+A)^2 - (k+A)^2\left(\frac{k}{a_1+1}\right)^h\right] \\
& + \sum_{k=1}^{a_2}\left[(k-A)^2 - (k-A)^2\left(\frac{k}{a_2+1}\right)^h\right]\Bigg\}.
\end{aligned}
$$

Expanding $(k+A)^2 = A^2 + 2Ak + k^2$ and $(k-A)^2 = A^2 - 2Ak + k^2$ and using the definition of $D = D(a_1, a_2, h)$ in Definition 3.1, one has

$$
\begin{aligned}
Var(Y) = & \frac{1}{D}\Bigg\{A^2 D + \sum_{k=1}^{a_1}\left[2Ak + k^2 - 2Ak\left(\frac{k}{a_1+1}\right)^h - k^2\left(\frac{k}{a_1+1}\right)^h\right] \\
& + \sum_{k=1}^{a_2}\left[-2Ak + k^2 + 2Ak\left(\frac{k}{a_2+1}\right)^h - k^2\left(\frac{k}{a_2+1}\right)^h\right]\Bigg\} \\
= & \frac{1}{D}\left\{A^2 D - 2A^2 D + \sum_{k=1}^{a_1}\left[k^2 - k^2\left(\frac{k}{a_1+1}\right)^h\right] + \sum_{k=1}^{a_2}\left[k^2 - k^2\left(\frac{k}{a_2+1}\right)^h\right]\right\}
\end{aligned}
$$

that yields the result.
(iii) and (iv) are obtained by considering the limits in h of individual probabilities $f(y;m,a_1,a_2,h)$. Indeed, it is easy to see that

$$
D(a_1,a_2,h) \to \begin{cases} (a_1+a_2+1) - a_1 - a_2 = 1 & \text{as } h \to 0 \\ (a_1+a_2+1) - 0 - 0 = (a_1+a_2+1) & \text{as } h \to \infty \end{cases}
$$

and finally

$$
f(y;m,a_1,a_2,h) \to \begin{cases} \mathbf{1}_{\{m\}}(y) & \text{as } h \to 0 \\ (a_1+a_2+1)^{-1}\,\mathbf{1}_{\aleph_{m,a_1,a_2}}(y) & \text{as } h \to \infty. \blacksquare \end{cases}
$$

4 General discrete triangular distribution

The most general extension of $\mathcal{DT}(m;a;h)$ can be defined as follows.

Definition 4.1 *Let* $(m, a_1, a_2, h_1, h_2) \in \mathbb{Z} \times \mathbb{N} \times \mathbb{N} \times \mathbb{R}_+ \times \mathbb{R}_+$. *A distribution* $\mathcal{DT}(m;a_1,a_2;h_1,h_2)$ *is said to be a* ***general discrete triangular*** *distribution with*

mode m*, left arm* a_1*, right arm* a_2*, left order* h_1 *and right order* h_2*, if its pmf is* $f(y; m, a_1, a_2, h_1, h_2) =$

$$\frac{1}{D(a_1, a_2, h_1, h_2)} \left\{ \left[1 - \left(\frac{m-y}{a_1+1} \right)^{h_1} \right] \mathbf{1}_{\aleph^*_{a_1,m}}(y) + \left[1 - \left(\frac{y-m}{a_2+1} \right)^{h_2} \right] \mathbf{1}_{\aleph_{m,a_2}}(y) \right\},$$

with

$$D(a_1, a_2, h_1, h_2) = (a_1 + a_2 + 1) - (a_1+1)^{-h_1} \sum_{k=1}^{a_1} k^{h_1} - (a_2+1)^{-h_2} \sum_{k=1}^{a_2} k^{h_2}.$$

Remark 4.2 *For* $h_1 = h_2 = h$ *we get the discrete (h-order) triangular distribution of Definition 3.1 as* $\mathcal{DT}(m; a_1, a_2; h, h) = \mathcal{DT}(m; a_1, a_2; h)$.

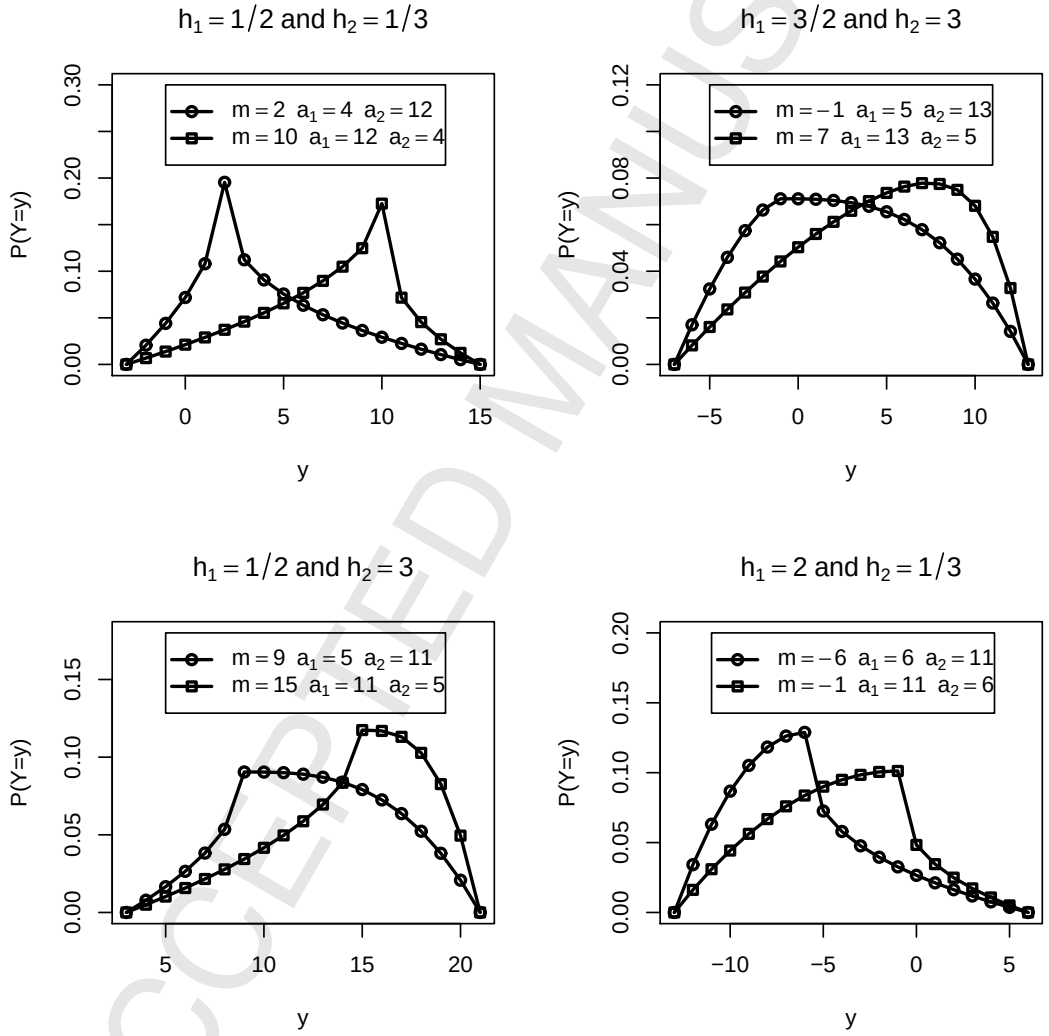


Fig. 2. Some general discrete triangular distributions.

Figure 2 illustrates different shapes that can be obtained by changing some parameters of $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$. The following theorem is similar to Theorem 3.3 and we omit its proof.

Theorem 4.3 *Let $Y \sim \mathcal{DT}(m; a_1, a_2; h_1, h_2)$. Then:*
(i) $\mathbb{E}(Y) = m + A(a_1, a_2; h_1, h_2)$ *with*

$$A(a_1, a_2; h_1, h_2) = \frac{1}{D(a_1, a_2, h_1, h_2)} \left[\frac{a_2 (a_2 + 1)}{2} - \frac{a_1 (a_1 + 1)}{2} + \sum_{k=1}^{a_1} k \left(\frac{k}{a_1 + 1} \right)^{h_1} - \sum_{k=1}^{a_2} k \left(\frac{k}{a_2 + 1} \right)^{h_2} \right];$$

(ii) $Var(Y) = B(a_1, a_2; h_1, h_2) - [A(a_1, a_2; h_1, h_2)]^2$ *does not depend on* m *with*

$$B(a_1, a_2; h_1, h_2) = \frac{1}{D(a_1, a_2, h_1, h_2)} \left[\frac{a_2 (a_2 + 1) (2a_2 + 1)}{6} + \frac{a_1 (a_1 + 1) (2a_1 + 1)}{6} - \sum_{k=1}^{a_1} k^2 \left(\frac{k}{a_1 + 1} \right)^{h_1} - \sum_{k=1}^{a_2} k^2 \left(\frac{k}{a_2 + 1} \right)^{h_2} \right];$$

(iii) when $h_1 \to 0$ *and* $h_2 \to 0$, $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ *tends to* $\mathcal{D}(m)$*;*
(iv) when $h_1 \to \infty$ *and* $h_2 \to \infty$, $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ *tends to* $\mathcal{U}(\aleph_{m,a_1,a_2})$*;*
(v) when $h_1 \to 0$ *and* $h_2 \to \infty$, $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ *tends to* $\mathcal{U}(\aleph_{m,a_2})$*;*
(vi) when $h_1 \to \infty$ *and* $h_2 \to 0$, $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ *tends to* $\mathcal{U}(\aleph_{a_1,m})$.

Remark 4.4 *The limits in* h_1 *and* h_2 *of mean* (i) *and variance* (ii) *of* $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ *give the same means and variances of the corresponding limit distribution* $(iii) - (vi)$ *of Theorem 4.3.*

5 Applications to discrete kernel estimators

Let $X_1, \cdots, X_n$ be independent and identically distributed (i.i.d.) discrete random variables with an unknown pmf p on the support $\mathbb{T}$. To simplify we shall assume $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \{0, 1, \cdots, N\}$ with known $N \in \mathbb{N}^*$. A discrete associated-kernel estimator $\widetilde{p}_n$ of p is defined as:

$$\widetilde{p}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T}, \tag{2}$$

where $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters (or bandwidths) that fulfills $\lim_{n \to \infty} h(n) = 0$, and $K_{x,h}$ is the "discrete associated-kernel" with the target x and the bandwidth h. Up to the normalizing constant $C = \sum_{x \in \mathbb{T}} \widetilde{p}_n(x)$, we assume that $x \mapsto \widetilde{p}_n(x)$ is a pmf. From Kokonendji *et al.* (2007) we have the following general definition.

Definition 5.1 *Let* $\mathbb{T}$ *be the discrete support of the pmf* p*, to be estimated,* x *a fixed target in* $\mathbb{T}$ *and* $h > 0$ *a bandwidth. A pmf* $K_{x,h}$ *on its support* $\mathbb{S}_x$ *(not depending on*

h) is said to be a ***discrete associated-kernel****, if it satisfies the following conditions:*

$$x \in \mathbb{S}_x, \tag{3}$$

$$\lim_{h\to 0} \mathbb{E}(Z_{x,h}) = x, \tag{4}$$

$$\lim_{h\to 0} Var(Z_{x,h}) = 0, \tag{5}$$

where $Z_{x,h}$ is the discrete random variable whose pmf is $K_{x,h}$.

In order to construct a discrete associated-kernel $K_{x,h}$ from a parametric discrete probability distribution K_θ, $\theta \in \Theta \subset \mathbb{R}^d$, on the support $\mathbb{S}_\theta$ such that $\mathbb{S}_\theta \cap \mathbb{T} \neq \emptyset$, we need to establish a correspondence between $(x, h) \in \mathbb{T} \times (0, \infty)$ and $\theta \in \Theta$. In what follows, we will call $K \equiv K_\theta$ the *type of discrete kernel* to make the difference with the classical notion of continuous kernels. In this context, the choice of the discrete associated-kernel becomes important as well as of the bandwidth. Also, given a type of discrete kernel K, the construction of any discrete associated-kernel is not obviously unique. As observed by Abdous and Kokonendji (2009), there are not many discrete associated-kernels in the sense of Definition 5.1; see, for example, Aitchison and Aitken (1976), Wang and van Ryzin (1981), Kokonendji *et al.* (2007).

Here we consider the discrete (h-order) triangular distribution $\mathcal{DT}(m; a_1, a_2; h)$ (see Definition 3.1) as a type of discrete kernel and, therefore, $\widetilde{p}_n$ of (2) becomes

$$\widetilde{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f\left(X_i; x, a_1, a_2, h\right), \quad x \in \mathbb{T}, \tag{6}$$

for fixed arms $(a_1, a_2) \in \mathbb{N}^2$. That is $K_{x,h}(\cdot) \equiv f(\cdot; x, a_1, a_2, h)$ which satisfies easily all conditions (3)-(5) with $\mathbb{S}_x = \aleph_{x,a_1,a_2}$ given in (1). Then, the pointwise bias can be expressed as

$$\begin{aligned} Bias\left[\widetilde{p}_n(x)\right] &= p\left[\mathbb{E}(Z_{x,a_1,a_2,h})\right] - p(x) + \frac{1}{2}Var\left(Z_{x,a_1,a_2,h}\right)p^{(2)}(x) + o(h) \\ &= A(a_1, a_2; h)p^{(1)}(x) + \frac{1}{2}\left\{B(a_1, a_2; h) - \left[A(a_1, a_2; h)\right]^2\right\}p^{(2)}(x) + o(h), \end{aligned} \tag{7}$$

where $Z_{x,a_1,a_2,h}$ is the random variable following $\mathcal{DT}(x; a_1, a_2; h)$ and $p^{(k)}$ is the finite difference of order $k \in \{1, 2\}$ (see, e.g., Kokonendji *et al* ., 2009). For the pointwise variance, we have

$$\begin{aligned} Var\left[\widetilde{p}_n(x)\right] &= \frac{1}{n}\, p(x)\left[1 - p(x)\right]\left[\Pr\left(Z_{x,a_1,a_2,h} = x\right)\right]^2 + R_n\left(x; a_1, a_2, h\right) \\ &= \frac{1}{n\left[D(a_1, a_2, h)\right]^2}\, p(x)\left[1 - p(x)\right] + R_n\left(x; a_1, a_2, h\right) \end{aligned}$$

with $R_n(x; a_1, a_2, h) \to 0$ when $n \to \infty$ and $h = h(n) \to 0$. In practice, both arms a_1 and a_2 are small and equal to 1, 2 or 3.

However, the general condition (3) can be replaced by $\bigcup_{x\in\mathbb{T}} \mathbb{S}_x \supseteq \mathbb{T}$. This means that the discrete associated-kernel takes into consideration the support $\mathbb{T}$ of the pmf p

to estimate. If $\bigcup_{x\in\mathbb{T}} \mathbb{S}_x$ is not equal to $\mathbb{T}$ (i.e. $\bigcup_{x\in\mathbb{T}} \mathbb{S}_x \supsetneq \mathbb{T}$) then one has a problem of boundary bias.

Assuming $\mathbb{T} = \mathbb{N}$ in (6), for fixed $a_1 \neq 0$, these discrete triangular kernel estimators induce a boundary bias on the left of $\mathbb{N}$ because the set $\bigcup_{x\in\mathbb{N}} \aleph_{x,a_1,a_2} = \{-a_1, \cdots, -1\} \cup \mathbb{N}$ contains strictly the support $\mathbb{N}$ of the unknown pmf p. In the symmetric case with $a_1 = a_2 = a$, Kokonendji *et al.* (2007, 2009a) have used an artificial modification of the arm a for significant observations to the left boundary $\{0, 1, \cdots, r\}$ (r too small, like 0, 1 or 2). For the present situation (6) of asymmetric discrete triangular kernel estimators, the solution is to consider the modified left arm a_0 of a_1 such that, for given $x \in \mathbb{N}$,

$$a_0 = \begin{cases} x & \text{if } \ x \in \{0, 1, \cdots, a_1 - 1\} \\ a_1 & \text{if } \ x \in \{a_1, a_1 + 1, \cdots, N, \cdots\}. \end{cases} \tag{8}$$

In particular, if $a_1 = a_2 = a$ we have asymmetric discrete associated-kernels for the left boundary $\{0, 1, \cdots, a - 1\}$ of $\mathbb{N}$ and symmetric ones on $\{a, a + 1, \cdots\}$ with $A(a, a, h) = 0$ for all $h > 0$ in (7); thus, it is different to the solution proposed by Kokonendji *et al.* (2007, 2009a), especially at the origin $x = 0$ of $\mathbb{N}$. The procedure (8) is so natural for asymmetric discrete triangular kernels and it is also more appropriate for smoothing any count distribution. Note that, by considering the solution (8) in the bias property (7), it is possible over weight the boundary; but, in practice, this procedure is done before the normalization of the estimator $\widetilde{p}_n$ by the constant $C = \sum_{x\in\mathbb{T}} \widetilde{p}_n(x)$ (e.g. Kokonendji *et al* ., 2007).

If $\mathbb{T} = \{0, 1, \cdots, N\}$ in (6) we must also take into account the boundary bias on the right of $\{0, 1, \cdots, N\}$ because of

$$\bigcup_{x\in\{0,1,\cdots,N\}} \aleph_{x,a_1,a_2} = \{-a_1, \cdots, -1\} \cup \{0, 1, \cdots, N\} \cup \{N + 1, \cdots, N + a_2\}.$$

Applying the modification of the left arm (8) on the right of $\{0, 1, \cdots, N\}$ (that is, at the neighbourhood of the point $x = N$), the modified right arm a_N of a_2 is such that, for given $x \in \{0, 1, \cdots, N\}$,

$$a_N = \begin{cases} a_2 & \text{if } \ x \in \{0, 1, \cdots, N - a_2\} \\ N - x & \text{if } \ x \in \{N - a_2 + 1, \cdots, N - 1, N\}. \end{cases} \tag{9}$$

Combining (8) and (9), these modified (asymmetric) discrete triangular kernel estimators are more appropriate for any compact pmf (or discrete functions) and also possibly for ordered categorial distribution (e.g. Aitchison and Aitken, 1976). They are flexible and offer a lot of possibilities than the (modified) symmetric discrete triangular kernel estimators (e.g. Kokonendji *et al.*, 2007), which are particular cases. All theoretical calculations (e.g. mean integrated squared error) and the usefulness of new models in practical count data analysis can be done like in Kokonendji *et al.* (2007, 2009a, 2009b) and we here omit them. Another way of these straightforward extensions of discrete triangular distributions would be the multivariate case and their multiple applications.

Acknowledgements. This work was partially supported by grants from FAPESP and UMR 6623 CNRS. We thank Gaston M. N'Guérékata for his careful reading along with the Associate Editor for his valuable comments.

References

Abdous, B. and Kokonendji, C.C. (2009). Consistency and asymptotic normality for discrete associated-kernel estimator. *African Diaspora J. Math.* **8**, 63-70.

Aitchison, J. and Aitken, C.G.G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* **63**, 413–420.

Johnson, D. (1997). The triangular distribution as a proxy for the beta distribution in risk analysis. *The Statistician* **46**, 387-398.

Johnson, N.L., Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions* (3rd ed.). John Wiley & Sons, New York.

Kokonendji, C.C., Senga Kiessé, T. and Zocchi, S.S. (2007). Discrete triangular distributions and nonparametric estimation for probability mass function. *J. Nonparam. Statist.* **19**, 241-254.

Kokonendji, C.C., Senga Kiessé, T. and Balakrishnan, N. (2009a). Semiparametric estimation for count data through weighted distributions. *J. Statist. Plann. Inference* **139**, 3625-3638.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009b). Appropriate kernel regression on a count explanatory variable and applications. *Adv. Appl. Statist.* **12**, 99-126.

Wang, M.-C. and Van Ryzin, J. (1981). A class of smooth estimators for discrete distributions. *Biometrika* **68**, 301-309.